Analysis of English Language Achievement Among Wisconsin English Language Learners

Prepared for the Wisconsin Department of Public Instruction

By
Christopher Babal
Shiyao Cao
Angela Filer
Isaac Hedtke
Keo Lo

Workshop in Public Affairs
Spring 2015
Table of Contents

List of Tables ................................................................................................................................. vi
List of Figures ................................................................................................................................ vi
Foreword ....................................................................................................................................... vii
Acknowledgments ........................................................................................................................ viii
Executive Summary ....................................................................................................................... ix
Introduction ..................................................................................................................................... 1
  ELL Programs ............................................................................................................................. 1
  Achievement Gap ....................................................................................................................... 2
Current Reporting on ELL Achievement ........................................................................................ 3
  Annual Measurable Achievement Objectives ............................................................................. 4
  AMAO Accountability ................................................................................................................ 4
  Limitations of Current Approach ............................................................................................ 5
  ACCESS Exams .......................................................................................................................... 5
Factors Influencing Performance .................................................................................................... 7
Data Description ............................................................................................................................. 9
Results on ACCESS Exam ........................................................................................................... 12
  Results ....................................................................................................................................... 13
  Language Groups ..................................................................................................................... 14
  Low Income Students ............................................................................................................... 14
  Disability Status ....................................................................................................................... 14
  Key Independent Variables: Age and Initial Proficiency ......................................................... 15
  Robustness Tests ..................................................................................................................... 17
Results on Success in Reaching English Proficiency and on Duration ........................................ 17
  Proficiency Model ................................................................................................................... 18
  Duration Model ....................................................................................................................... 19
  Robustness Tests ..................................................................................................................... 19
Discussion ..................................................................................................................................... 19
  Limitations ................................................................................................................................ 21
Recommendations ......................................................................................................................... 21
Conclusion .................................................................................................................................... 22
Appendix A: Dual-Language Programs ........................................................................................ 24
Appendix B: AMAO Reports ....................................................................................................... 25
Appendix C: Federal Funding ...................................................................................................... 34
Appendix D: English Language Proficiency Levels ..................................................................... 35
Appendix E: Uses of Raw, Scale, and Proficiency Level Scores ................................................. 37
Appendix F: Student Growth Percentiles ..................................................................................... 38
References ..................................................................................................................................... 39
List of Tables

Table 1. Translation of ACCESS Scale Scores to Composite Scores ............................................ 7
Table 2. Summary of Variables .................................................................................................... 10
Table 3. Detail of DPI Dataset Demographics and Authors’ Models........................................... 12
Table 4. Associations of Student Characteristics on Subsequent Test Scores......................... 14
Table 5. Effect Sizes of Changes from Initial Proficiency in Standard Deviations...................... 16
Table 6. Results: Proficiency and Duration .................................................................................. 18

List of Figures

Figure 1. Percentage of Students Below Basic in Math and Reading Nationally, in 2005 ............ 3
Figure 2. Number of Students Starting ELL Programs by Grade.................................................. 10
Figure 3. Number of Students in Sample with Key Information.................................................. 11
Figure 4. Grade Effect on ACCESS Subtest Scores..................................................................... 15
Foreword

This report is the result of collaboration between the La Follette School of Public Affairs at the University of Wisconsin–Madison and Wisconsin Department of Public Instruction. Our objective is to provide graduate students at La Follette the opportunity to improve their policy analysis skills while contributing to the capacity of Wisconsin education policymakers to understand how students learning English are faring.

The La Follette School offers a two-year graduate program leading to a master’s degree in public affairs. Students study policy analysis and public management, and they can choose to pursue a concentration in a policy focus area. They spend the first year and a half of the program taking courses in which they develop the expertise needed to analyze public policies. The authors of this report are all in their final semester of their degree program and are enrolled in Public Affairs 869 Workshop in Public Affairs. Although acquiring a set of policy analysis skills is important, there is no substitute for doing policy analysis as a means of learning policy analysis. Public Affairs 869 gives graduate students that opportunity.

This year the workshop students were divided into eight teams. Other teams completed projects for the City of Madison, the Wisconsin Department of Children and Families, the Wisconsin Legislative Council, Madison Metro Transit System, the Center for Economic Progress, and The Financial Clinic of New York City.

The report builds on research that shows an achievement gap between English language learners and other students. How can these students be helped? Current techniques for assessing the progress of these students offer relatively limited insight into how educators might help these students. Unrealistic federal mandates will soon result in labeling all Wisconsin schools as failing, failing to provide any insight into variation for why some students fare better than others. This report offers some alternative measures of progress of English language learners, as well as some insights into what factors make a difference in that progress.

Donald Moynihan
Professor of Public Affairs
Madison, Wisconsin
May 2015
Acknowledgments

We thank Jared Knowles and Carl Frederick of the Wisconsin Department of Public Instruction for providing background information, feedback, and support throughout our research and analysis. We are grateful for the helpful insights of WIDA Research Director Dr. Gary Cook, and Audrey Lesondak and Phil Olsen of the Wisconsin Department of Public Instruction for sharing their ideas. We appreciate the invaluable assistance of Mark Banghart and the dedicated staff at the Social Science Computing Cooperative. Finally, we extend our gratitude to Professor Donald Moynihan for his guidance and advice, and Karen Faster for her editorial assistance and expertise.
Executive Summary

An academic achievement gap exists between English language learners (ELLs) and non-ELLs nationally and within Wisconsin K-12 schools. Wisconsin public schools serve more than 47,000 ELL students. ELL students make up 5 percent of the public school population, and this will continue to increase in the coming years.

The existing approach to assessing the progress of ELLs is limited in a number of respects. Assessments do not adequately detail strengths and weaknesses of ELL programs and the ELL population. Schools receive reports that detail ELL students’ continued improvement in terms of adequate yearly progress, reaching English proficiency, test participation rates, and graduation rates. The Wisconsin Department of Public Instruction (DPI) seeks the construction of metrics that help better explain the performance of ELLs given students’ initial age and initial English proficiency upon entering an ELL program, measured in terms of time to English proficiency.

Using data from the DPI Longitudinal Data System, we focus on the effect size of starting age and initial English proficiency on later performance, probability of reaching proficiency, and expected duration in an ELL program. We find that an increase in initial English proficiency by one standard deviation is positively correlated with an increase in later test scores by 0.4 standard deviations. An increase in initial proficiency also increases the probability of reaching proficiency and reduces the expected time spent in an ELL program by greater than eight months. As starting age increases, students demonstrate a positive impact on later scores until fifth grade, at which time the effect becomes negative. As age increases, the probability of reaching proficiency decreases; however, the expected duration in an ELL program is shortened by more than two months. Student performance varies by language group. Spanish- and Hmong-speaking students, the two largest minority language groups in Wisconsin, demonstrated slower improvement rates compared to other language groups. Consistent with prior work we found that poorer and disabled students fare less well than other students on the Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS) exam Wisconsin uses to measure English proficiency.

We make four recommendations to DPI to facilitate local administrative understanding of their ELL students:

- First, we recommend that DPI provide reports to school districts that separately analyze student ACCESS exam scores in reading, writing, listening, and speaking.
- Second, to school districts with large ELL populations, we recommend that DPI provide reports that analyze test scores by subgroups within the district’s ELL population.
- Third, to help evaluate effectiveness of ELL programs, we recommend that DPI measure student growth percentiles, which compare students to cohorts with similar characteristics and test performance, and report percentiles to school districts as well as to ELL students.
- Finally, we also recommend that DPI explore the WIDA Consortium’s new dashboard that provides online visualizations of ELL statistics and create a short user manual for district administrators.
Introduction

National standardized test data shows that an academic achievement gap exists between English language learners (ELLs) and non-ELLs. Seventy-one percent of eighth-grade ELL students scored below basic in reading, compared to 19 percent of white-non ELL students nationally (Fry 2007). The achievement gaps between ELL and non-ELL students are especially important given the projected growth of the two major language groups in Wisconsin, Spanish and Hmong. From 2000 to 2010, the Hmong-speaking population in Wisconsin increased by 46 percent (Pabst 2013). Nationally, students who speak a language other than English at home and have varying levels of English proficiency are anticipated to make up 40 percent of all elementary and secondary students by 2030 (Thomas & Collier 2003).

The current system of reporting ELL achievement, established under No Child Left Behind, is limited in the information it provides. Because of the increasingly unrealistic performance goals established by the No Child Left Behind Act of 2001 within the structure of these reports, all public schools in Wisconsin will be classified as failing the federal law’s criteria within the next school year. Such stark categorizations do little to provide an evidence-based understanding of variation in progress across schools, school districts, or groups of students.

School districts look to the Wisconsin Department of Public Instruction (DPI) for leadership and further detail about the strengths and weaknesses within their ELL programs, particularly factors that influence student performance. In proposing this project, DPI asked us to examine how students’ starting age and initial English proficiency level influence student performance on English language assessments. Districts cannot control these factors, but starting age and initial proficiency do affect students’ success.

We explore the following question in the context of Wisconsin K-12 schools: How do starting grade in an ELL program and initial English proficiency level affect a student’s achievement in the ELL program? Better understanding the answer to this question will allow us to recommend more effective metrics and reporting techniques to improve information about ELL programs at the state and district levels. School districts may use this information to better assess the strengths and weaknesses of ELL programs and inform decisions about how they target resources within the district.

We use a variety of approaches to measure student achievement: annual English language assessment scores, student duration in an ELL program, and the probability of achieving English proficiency while in school.

This report begins by discussing ELL programs, the ELL achievement gap, and current reporting on ELL achievement. We explore factors that influence ELL student achievement and then describe the data we use in our analysis of ELL achievement, the models we constructed, and the results we found. We end our analysis by discussing the implications of these results and providing recommendations to DPI.

ELL Programs

The No Child Left Behind Act categorizes students acquiring English for their education as being of “limited English proficient” (Wisconsin DPI 2015c). The federal government also refers
to these students as English as a second language students or bilingual students. Most recently, the state of Wisconsin refers to these students as ELLs.

ELL programs help students become bilingual and biliterate. The federal government, through the Fourteenth Amendment to the Constitution as well as several Supreme Court rulings,\(^1\) established the rights of ELL students and defined the legal responsibilities of school districts serving ELL students. DPI defines the purpose of ELL programs as “to assist school districts serving ELL students to become proficient in academic English and prepared for their continuing education and career readiness goals,” (Wisconsin DPI 2015c). State and federal law require that ELL students attending public schools must receive the opportunity to participate in and benefit from educational programming in a meaningful way (Wisconsin DPI 2015e).

To determine if a program meets the needs of ELL students, a federal appeals court ruled in Castaneda v. Pickard (1981) that administrators must consider the following questions:

1) Has the school designed a program based on a sound educational theory?
2) Has the school pursued its program with adequate resources, personnel, and practices?
3) Has the program achieved satisfactory results? (Casta v. Pickard, 1981; 5th Circuit US Court of Appeals).

DPI identifies four types of dual-language programs in Wisconsin. They differ mainly in the population of students served: developmental bilingual, two-way immersion, heritage language, and world language immersion programs (Wisconsin DPI 2015d). For a more detailed description of dual-language programs, see Appendix A.

**Achievement Gap**

Research has established that an achievement gap exists between ELL and non-ELL students. The National Assessment of Educational Progress, which contains national standardized testing data and acts as the “Nation’s Report Card,” shows an achievement gap exists between ELL students and non-ELL students. The assessment measures student performance at grades four and eight in reading and mathematics. Four achievement levels categorize performance on the assessment: below basic, basic, proficient, and advanced.

Research by the Pew Hispanic Center identified the achievement gap between ELL and non-ELL students as the differences in percentages at or above the basic level for ELL students and a comparison group. Put another way, this measures how far ELL students lag in at least partial mastery of prerequisite skills. The 2005 assessment found that 46 percent of ELL students nationwide scored below the basic level in mathematics. In reading, 77 percent of ELL fourth-graders scored below the basic level. Among non-ELL fourth-graders, 11 percent scored below the basic level in mathematics, and 25 percent scored below the basic level in reading (Fry 2007).

---

\(^1\) Other important federal court rulings that addressed ELL education include Plyler v. Doe, 1982, which ruled that the Fourteenth Amendment prohibits states from denying a free public education to undocumented immigrant children regardless of their immigrant status, and Lau v. Nichols, 1974, which the Equal Educational Opportunities Act of 1974 codified into federal law.
The ELL achievement gap widens at higher grades. The 2005 National Assessment of Educational Progress assessment found that the gap in achievement in fourth-grade math is 35 points (89 percent at or above basic level for whites versus 54 percent for ELL students). The eighth-grade exam revealed a 50-point gap between ELL and non-ELL students (Fry 2007). From grades three to eight, the achievement gap in Wisconsin ranged from 35 to 40 points in mathematics and 35 to 42 points in reading. Figure 1 shows the percentage of students scoring below basic in math and reading on this exam.

![Figure 1. Percentage of Students Below Basic in Math and Reading Nationally, in 2005](image)

Wisconsin public schools in 2012 served more than 47,000 ELL students, which amounts to more than 5 percent of the state’s public school population. More than 80 percent of school districts in Wisconsin have ELL students (Wisconsin DPI 2015c). Population trends suggest that Wisconsin will mirror national trends in seeing increased numbers of ELLs (Thomas & Collier 2003). Since such a high number of school districts have ELLs and the number of ELLs served by public schools in Wisconsin will increase, learning more about ELL achievement gaps is an important goal for DPI. In addition, DPI has identified ELL achievement gaps as a policy concern requiring immediate attention; therefore, the agency seeks the development of metrics that help explain the performance of ELL students in the state and understanding of the role of initial English proficiency and initial starting age upon entering an ELL program.

**Current Reporting on ELL Achievement**

The No Child Left Behind Act mandates that states test all students in reading and mathematics in grades three through eight, and again in high school. Student performance on these exams determines school, district, and state accountability scores and assesses academic progress.

No Child Left Behind Act also enacted major changes to the expectations placed on state and local agencies regarding accountability for several subgroups of students, including low income students, students with disabilities, students in major racial and ethnic groups, and students with
limited English proficiency. Specifically, Title III of the act requires all states to do the following:

- establish English language proficiency standards aligned to state academic content standards, yet suitable for students learning English as a second language;
- annually assess the English language proficiency of each ELL student, using a valid and reliable assessment aligned with federal and state standards;
- define annual measurable achievement objectives to report on progress toward and attainment of English proficiency and academic achievement standards; and
- hold local educational agencies accountable for meeting increasing Annual Measurable Achievement Objective (AMAO) targets for English language proficiency over time, (No Child Left Behind 2002, Public Law 107-110, 115 Statute 1425).

Annual Measurable Achievement Objectives

Title III of the No Child Left Behind Act establishes three specific AMAOs:

1. Progressing in English language acquisition: annual increases in the number or percentage of students making progress in learning English;
2. Reaching English language proficiency: annual increases in the number or percentage of students attaining English language proficiency by the end of each school year;
3. ELL AMAO: adequate progress for ELLs in grade-level academic achievement standards for reading and mathematics, test participation, and high school graduation (Wisconsin DPI 2015a).

Wisconsin uses the Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS) exam Wisconsin uses to measure English proficiency. Wisconsin school districts receive AMAO reports based on those test results to help evaluate the effectiveness of the district’s ELL programs. The reports outline the three AMAOs and whether that district met each of them, and they provide visualizations of the data (Wisconsin DPI 2014a). See Appendix B for details about the AMAO reports that school districts receive from DPI. The AMAO reports do not include detail on all information included in the ACCESS exams, such as trends in the subtests of reading, writing, listening, and speaking.

AMAO Accountability

If a school district does not fulfill a certain AMAO, DPI takes action, and districts can face a range of consequences. The first year a district does not meet an AMAO, it must notify parents of ELL students within 30 days. DPI will provide technical assistance to the district as needed.

If a district fails to meet an AMAO for a second consecutive year, in addition to parental notification, the district must develop and adopt an improvement plan that addresses specific factors that prevented the district from meeting the AMAO. The district must utilize professional development strategies and activities developed in consultation with DPI. DPI provides technical assistance to develop instructional strategies. If a district fails to meet an AMAO for a third consecutive year, procedures for the school district and DPI repeat.

In the event of four consecutive failures to fulfill an AMAO, the district must modify the curriculum, program, and method of instruction or replace personnel, at which time DPI
determines whether the district will continue to receive funds related to the failure to meet AMAOs (Wisconsin DPI 2014b). For more information about federal funding for ELL programs, see Appendix C.

**Limitations of Current Approach**

School districts have asked DPI for AMAO reports that tell them more about their ELL programs. The current approach has a number of shortcomings.

*Unrealistic expectations:* The No Child Left Behind Act established goals of Adequate Yearly Progress that schools and districts should aim to achieve, as measured by standardized tests. The act delegated to the states the task of establishing yearly benchmarks. Many states chose smaller increments of improvement in the early years of enactment to show that they met their Adequate Yearly Progress goals (Usher 2011). This standard meant that in the later years of the No Child Left Behind target date, districts need to show unrealistic levels of improvement. Since school districts will all eventually fail to reach Annual Yearly Progress objectives, districts seek more constructive feedback in reports from DPI.

*Lack of actionable information:* AMAO reports indicate to school districts the percentage of students who take exams and the graduate rate of ELL students. While this reporting does fulfill a federal mandate, it does not provide school districts with details about their ELL programs that can help improve outcomes for ELL students. Research demonstrates that starting age and initial English proficiency are important factors in determining success, measured in terms of testing out of ELL services (Hakuta, Butler & Witt 2000; Suárez-Orozco, Suárez-Orozco, & Todorova 2009). Current AMAO reports do not provide school districts with predictions or expectations of student duration in an ELL program.

*Lack of information about progress on different skills:* ACCESS exams cover four subtests: reading, writing, listening, and speaking. Current AMAO reports do not include analysis of these elements; rather, as mentioned, they simply indicate whether a school district has reached adequate yearly progress. Subtest analysis can help identify elements of the ELL program school districts can improve.

*Lack of information on subgroups:* The current approach does not allow for detailed subgroup analysis of ELL students. While Spanish and Hmong are the two largest language groups in Wisconsin, but more than 137 languages are spoken in the state (Wisconsin DPI 2015c). Language groups may perform differently on separate sections of the ACCESS exam. A more detailed look at different language groups can help school districts identify strengths and weaknesses of their ELL programs and direct resources toward improvement where necessary.

**ACCESS Exams**

Standardized tests evaluate students’ English proficiency levels. Wisconsin participates in the WIDA\(^2\) Consortium for testing ELLs. Students who qualify for ELL services take the WIDA-
ACCESS (WIDA 2014a) Placement Test before they enter ELL programs to determine if they need English language instructional services, and at what level. ACCESS is short for the Accessing Comprehension and Communication in English State-to-State for English Language Learners, the exam most Wisconsin school districts use to measure English language proficiency.

Students then take a separate exam, the ACCESS for ELLs exam, once every school year until they test out of ELL services (Abedi 2007). WIDA tailors both exams to each state’s federal and state testing requirements (Cook et al. 2008). Our analysis focuses on the ACCESS for ELLs exam, which allows for the evaluation of student performance over multiple years. For clarity, this report uses “ACCESS” to refer to the “ACCESS for ELLs” exam.

Wisconsin administrative code requires classifying each ELL on a scale from one to six:

1. Beginning Preproduction;
2. Beginning Production;
3. Intermediate;
4. Advanced Intermediate;
5. Advanced, and
6. Formerly Limited English Proficient and Now Fully English Proficient

Appendix D details the statutory definitions of these categories. The Wisconsin Department of Public Instruction adds a seventh category: Fully English Proficient and Never Limited English Proficient (Wisconsin DPI 2015f).

Wisconsin ELLs take the ACCESS exam once every school year until they achieve a score of 6 (Now Fully English Proficient), and exit ELL services. WIDA develops ACCESS exams for five grade clusters: kindergarten, grades 1 to 2, 3 to 5, 6 to 8, and 9 to 12. Except for kindergarten, each grade cluster breaks into overlapping tiers of beginning, intermediate, and advanced (WIDA 2014a).

Each ACCESS exam consists of four subtests to evaluate students’ ability to read, write, listen, and speak in English. Multiple-choice questions evaluate listening and reading. ACCESS evaluates writing using written tasks scored by a central testing company using a writing rubric. ACCESS evaluates speaking through scripted in-person interviews scored by the local test administrator (WIDA 2014a).

WIDA tabulates ACCESS exams using raw scores, scale scores, and English language proficiency levels. Raw scores calculate the number of correct responses out of the total number of possible responses. Scale scores weigh raw scores by the difficulty of each task. This weighing means that responding correctly to more difficult tasks is worth more than responding correctly to simpler tasks. Analysts can compare scale scores across grades 1 through 12, but cannot compare scale scores across subtests of reading, writing, listening, and speaking (WIDA 2014a).

Consortium no longer uses any acronym definition (WIDA Consortium 2015).
English language proficiency levels correspond to the statutory reporting scale of 1 through 6. WIDA calculates English language proficiency levels using scale scores at each grade level. A student will have a different English language proficiency level corresponding to each subtest score for reading, writing, listening, and speaking. WIDA also calculates English language proficiency levels for composites of scale scores: oral language, literacy, comprehension, and overall (WIDA 2014a). Table 1 outlines how scale scores translate into composite scores, using weighted percentages. Appendix E outlines the appropriate uses of raw, scale, and proficiency level scores.

<table>
<thead>
<tr>
<th>Oral Language</th>
<th>Listening</th>
<th>Speaking</th>
<th>Reading</th>
<th>Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literacy</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Comprehension</td>
<td>30%</td>
<td>70%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>15%</td>
<td>15%</td>
<td>35%</td>
<td>35%</td>
</tr>
</tbody>
</table>

Source: WIDA 2014a.

Factors Influencing Performance

Many factors influence ELL students’ ability to reach English proficiency, and some overlap with factors that affect student educational performance generally, including students’ low income status and enrollment in high-poverty schools. Some factors are specific to the challenges of learning English, including age of entry into an ELL program and initial English proficiency level.

Low income status increases difficulties for students learning English (Snow, Burns, & Griffin 1998). The majority of the ELL student population comes from low income families. Income affects students’ access to educational resources at both individual and school levels. Without resources such as reading materials, tutoring, and access to educational equipment such as computers puts students at a disadvantage. For example, students whose parents have a high education level and a high income have higher levels of English proficiency than students whose parents have a low education level or a low income (Bialystok & Hakuta 1994). A study by Kieffer (2010) found that low income ELL students experienced more difficulties at the developmental period of late elementary and middle school grades, compared to native English speakers. More specifically, low income ELL students tend to demonstrate a higher risk for developing reading difficulties during the same periods in late elementary and middle school grades.

All students require access to good schools for success. High poverty schools often have lower credential teachers, which affect all students, including ELL students. High poverty schools also are often racially segregated. Suárez-Orozco, Suárez-Orozco, and Todorova (2009) linked poverty and level of racial segregation to lower academic achievement of non-native English speaking students. Most ELL students in the United States attend high poverty schools with large percentages of ELL students limiting ELL students’ interaction with native English speakers (Liu et al. 1999).
Age of entry and initial English proficiency may significantly influence the time it takes for ELL students to reach English proficiency. The critical period hypothesis states that the developmental period when exposure to the language begins significantly influences an individual’s ability to gain full proficiency (Johnson & Newport 1989). Exposure to English at an earlier age is likely to enable ELL students to gain proficiency faster than if they began at a later age.

Though many studies demonstrate the effects of age on learning a new language, few studies examine the effect of age on the rate that ELL students reach English proficiency. Research suggests that children who start acquiring a language at a younger age fare better than older children, implying that students who enter into ELL programs earlier should perform better. However, no longitudinal studies with large populations have examined the correlation. The closest is a study done by Hakuta, Butler, and Witt (2000), which found that 90 percent of ELL students reach oral English proficiency within five years if they enter an ELL program in kindergarten. That study does not evaluate a large population of older ELL students beyond kindergarten. It also only evaluates oral English proficiency, not evaluate reading or writing.

Furthermore, age of entry may influence ELL students’ ability to reach English proficiency. If students enter an ELL program in kindergarten, and their test scores indicate that they are strong in listening and reading, but struggle in writing and speaking, teachers can focus more on the students’ weakness in the upcoming years. On the other hand, older students would have less of an opportunity to identify their weakness and overall less time for schools to help them overcome it. These scenarios again suggest that the younger students are when they enter ELL programs, the faster they should acquire English. Yet, because we are interested in not just acquisition of the English language, but also the how ELL students perform on the ACCESS exam, which is also influenced by students’ test taking ability, starting age may have a different effect on language acquisition than what prior research has shown.

Like age, student’s initial English proficiency level may affect students’ rates of learning English. A five-year longitudinal study of immigrant students explored the major factors affecting academic performance of ELL students. The study ranked initial English language proficiency and behavioral engagement (school problems and violence, relational engagement, cognitive engagement, and gender) as the two most influential variables that affect academic performance (Suárez-Orozco, Suárez-Orozco, & Todorova 2009).

To the best of our knowledge, scholars have not conducted substantial research on how initial English proficiency affects the performance of students learning English. The lack of studies results from a lack of longitudinal data that tracks large samples of ELL students for every year of an ELL program. It is important to investigate initial proficiency because students with higher proficiency levels have less of a proficiency gap compared to students with lower initial English language proficiency. This observation is critical because no matter what age students enter an ELL program, students with more English proficiency will need less improvement to test out compared to students with lower levels of English proficiency. This observation could lead to very different paths for students with varying levels of English proficiency. Additionally, students with higher levels of English may reach proficiency faster because they can better utilize and understand the resources they are provided.
The structure of an ELL program likely influences student performance. Though some literature suggests that ELL students perform better when taught in dual languages (Collier 2004), experts debate this idea. Quality of education also influences ELL students’ ability to reach English proficiency. Schools have different resources, and teacher quality can differ greatly. Although quality of education and access to dual language instruction may influence student performance, our dataset does not contain this information.

Data Description

Our analysis uses a DPI longitudinal dataset that includes Wisconsin public school students enrolled in an ELL program for the school years 2005-06 through 2013-2014. The dataset includes students who entered ELL programs before and during the years for which we have data, as well as students who exited ELL programs during and after the years for which we have data. This situation means that, for many students, we do not have data for every year that they were in an ELL program.

The data include annual test scores from the Wisconsin ACCESS for ELLs exam. The scores allow us to analyze trends in reading, writing, listening, and speaking. These data should provide more descriptive information for school districts to use when evaluating their programs. Additionally, ACCESS exams provide data for each year of the program.

We considered several other variables in the dataset in our analysis. Because income and disability status affect students’ academic performance, we examine the role of demographic variables in our models. Instead of using students’ age directly, we approach ELL program starting age by using their grade level when they first take the ACCESS exam. We decided to use grade rather than age because the reporting for grade is more complete in the dataset. The native language of the student measures both language and race, and we mainly focus on Spanish and Hmong speakers, indicated by a student’s native language. We measure students’ income by registration in free and reduced-price school lunch programs. We use individualized education plans to identify students with disabilities. Table 2 outlines the variables used in our regression models.
Table 2. Summary of Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Statistical Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Later subtest score</td>
<td>Average subtest standardized score of 2&lt;sup&gt;nd&lt;/sup&gt;, 3&lt;sup&gt;rd&lt;/sup&gt;, and 4&lt;sup&gt;th&lt;/sup&gt; ACCESS test</td>
<td></td>
</tr>
<tr>
<td>Probability of reaching proficiency</td>
<td>Whether an ELL reaches proficiency in our data set. If an ELL reaches 6 proficiency score, the student will test out the program</td>
<td></td>
</tr>
<tr>
<td>Duration</td>
<td>Number of years that ELL spend in the program to reach proficiency</td>
<td>Model 3: 5.4 year</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Min: 0 year</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max: 7 years</td>
</tr>
<tr>
<td>Starting age</td>
<td>Grade level of an ELL when entering the program</td>
<td>Model 1: 2.2&lt;sup&gt;th&lt;/sup&gt; grade</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kindergarten</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12&lt;sup&gt;th&lt;/sup&gt; grade</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Model 2: 1.6&lt;sup&gt;th&lt;/sup&gt; grade</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kindergarten</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6&lt;sup&gt;th&lt;/sup&gt; grade</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Model 3: 1.3&lt;sup&gt;th&lt;/sup&gt; grade</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kindergarten</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6&lt;sup&gt;th&lt;/sup&gt; grade</td>
</tr>
<tr>
<td>Square of starting age</td>
<td>Quadratic term for grade level</td>
<td></td>
</tr>
<tr>
<td>Initial subtest score of ACCESS test</td>
<td>Standardized score by grade level</td>
<td></td>
</tr>
<tr>
<td>Square of initial subtest score of ACCESS test</td>
<td>Quadratic term for initial subtest score of ACCESS test</td>
<td></td>
</tr>
<tr>
<td>Initial composite standardized score of ACCESS test</td>
<td>This score includes four components: listening, speaking, reading, and writing. We also standardize the composite score</td>
<td></td>
</tr>
<tr>
<td>Hmong</td>
<td>ELL students whose native language is Hmong.</td>
<td>Hmong will be 1, and other ELLs who speak neither Hmong nor Hispanic will be 0</td>
</tr>
<tr>
<td>Spanish</td>
<td>ELL students whose native language is Spanish.</td>
<td>Spanish will be 1, and other ELLs who speak neither Hmong nor Hispanic will be 0</td>
</tr>
<tr>
<td>Low-income students</td>
<td>Students who register in a free and reduced lunch program.</td>
<td>Low income status will be 1 for the dummy variable</td>
</tr>
<tr>
<td>Disabled students</td>
<td>Students who have Individual Education Program</td>
<td>Students has recognized disability will be 1 for the dummy variable</td>
</tr>
</tbody>
</table>

Source: Authors, using Wisconsin DPI ELL student dataset, 2005-06 through 2013-14.

Figure 2 shows that a large number of students enter ELL programs in kindergarten, with decreasing numbers each year through 12<sup>th</sup> grade.

**Figure 2. Number of Students Starting ELL Programs by Grade**

Source: Authors, using Wisconsin DPI ELL student dataset, 2005-06 through 2013-14.

We only include students in our analysis if the dataset contains information about the student’s initial proficiency and age of entering an ELL program, so that we can investigate the effects of these two variables. We could not include students who had entered ELL programs prior to 2005,
the first year of our dataset. Our analysis still includes a large number of students due to the longitudinal nature of the dataset. The reason for exclusion means that the missing data should not bias the results in any way. The dataset includes all of the data we require for our analysis for a total of 32,243 students. Figure 3 shows the number of students in the dataset by the type of information the dataset contains about each student. From top to bottom, the characteristics compound onto one another. For example, 110,542 students have grade level data and no known data errors.

<table>
<thead>
<tr>
<th>Total Students in Sample</th>
<th>110,711</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Known Errors</td>
<td>110,542</td>
</tr>
<tr>
<td>Grade Level Data</td>
<td>110,542</td>
</tr>
<tr>
<td>First Year Data</td>
<td>49,392</td>
</tr>
<tr>
<td>FRL Data</td>
<td>46,895</td>
</tr>
<tr>
<td>Test Scores</td>
<td>32,243</td>
</tr>
</tbody>
</table>

Source: Authors, using Wisconsin DPI ELL student dataset, 2005-06 through 2013-14.

Comparing demographic characteristics allows us to ensure each subgroup in our analysis adequately represents the population as a whole. Table 3 details the demographics for the dataset and students in each of our models. The sample sizes for proficiency and duration models include students who enter ELL programs in grades kindergarten through sixth grade. We made this decision to ensure that students in the sample had at least seven years to reach proficiency or graduate high school. Within our dataset, the average time to proficiency is just more than five years.
Table 3. Detail of DPI Dataset Demographics and Authors’ Models

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Total Data N=110,542</th>
<th>Later Score Model N=32,223-32,274</th>
<th>Proficiency Models N=5823-5823</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>42%</td>
<td>48%</td>
<td>52%</td>
</tr>
<tr>
<td>Male</td>
<td>46%</td>
<td>51%</td>
<td>47%</td>
</tr>
<tr>
<td>Unknown</td>
<td>6%</td>
<td>10%</td>
<td>1%</td>
</tr>
<tr>
<td>Missing</td>
<td>6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>26%</td>
<td>24%</td>
<td>31%</td>
</tr>
<tr>
<td>Black</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>52%</td>
<td>64%</td>
<td>52%</td>
</tr>
<tr>
<td>Indian</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>White</td>
<td>7%</td>
<td>9%</td>
<td>14%</td>
</tr>
<tr>
<td>Missing</td>
<td>13%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hmong</td>
<td>16%</td>
<td>14%</td>
<td>13%</td>
</tr>
<tr>
<td>Spanish</td>
<td>49%</td>
<td>28%</td>
<td>48%</td>
</tr>
<tr>
<td>Other</td>
<td>35%</td>
<td>58%</td>
<td>39%</td>
</tr>
<tr>
<td>Individual Education Plan</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>8%</td>
<td>6%</td>
<td>3%</td>
</tr>
<tr>
<td>No</td>
<td>79%</td>
<td>93%</td>
<td>95%</td>
</tr>
<tr>
<td>Unidentified</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>Missing</td>
<td>13%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-Income</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free/Reduced Lunch</td>
<td>74%</td>
<td>71%</td>
<td>60%</td>
</tr>
<tr>
<td>Initial Proficiency Score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hmong-Speaking</td>
<td>3.06</td>
<td>2.45</td>
<td>2.80</td>
</tr>
<tr>
<td>Spanish-Speaking</td>
<td>2.63</td>
<td>2.42</td>
<td>2.97</td>
</tr>
<tr>
<td>Other</td>
<td>3.25</td>
<td>3.18</td>
<td>3.72</td>
</tr>
</tbody>
</table>

Source: Authors, using Wisconsin DPI ELL student dataset, 2005-06 through 2013-14.

Results on ACCESS Exam

We used regression analysis to investigate how starting age and initial English proficiency level affect students later academic achievement. Our models measure academic growth using standardized scores on the ACCESS exam. Our models do not measure academic achievement based scores on the Wisconsin Knowledge and Concepts Examination, because we are
comparing student achievement within the ELL population. For subgroup comparison, ACCESS scale scores offer more descriptive information.

The dependent variable in our models estimates students’ later ACCESS subtest scores. We use a student’s third ACCESS score as a dependent variable, testing how the student’s first test score and grade at entering an ELL program affect her/his third test score. Since some students might do worse on an individual test for idiosyncratic reasons, we averaged the second, third, and fourth test scores to approximate the third score as the dependent variable. We estimate baseline models with an ordinary least squares method.

The progression of this model is listed below. Subtests, whether for the later score or the initial score, refers to the four components of the ACCESS exam. These components are listening, speaking, reading, and writing. The progression of models was conducted for each subtest component. Refer to Table 2 for additional information on the variables.

The first regression examines how the later listening score relates to the two key independent variables of initial age and initial listening score, and controls for other variables.

\[
Later\ Subtest\ Score = \beta_0 + \beta_1 Starting\ Age + \beta_2 Initial\ Subtest\ Score + \beta_3 Hmong + \beta_4 Spanish + \beta_5 Low\ Income + \beta_6 Disabled + \epsilon
\]

The relationships among the dependent and independent variables might not be linear. Therefore, the next regression in model 1 includes a quadratic term for starting age to examine whether the model explains more of the effects of an older starting age on future performance on the ACCESS exam.

\[
Later\ Subtest\ Score = \beta_0 + \beta_1 Starting\ Age + \beta_2 Square\ of\ Starting\ Age + \beta_3 Initial\ Subtest\ Score + \beta_4 Hmong + \beta_5 Spanish + \beta_6 Low\ Income + \beta_7 Disabled + \epsilon
\]

The final regression for model one also included a quadratic term for the initial listening score to potentially explain more of the effects of initial proficiency.

\[
Later\ Subtest\ Score = \beta_0 + \beta_1 Starting\ Age + \beta_2 Square\ of\ Starting\ Age + \beta_3 Initial\ Subtest\ Score + \beta_4 Square\ of\ Initial\ Subtest\ Score + \beta_5 Hmong + \beta_6 Spanish + \beta_7 Low\ Income + \beta_8 Disabled + \epsilon
\]

Results

Table 4 summarizes the results in terms of standard deviations of ACCESS subtest scores. The results are explained in terms of comparable effect sizes. The interpretation of the results below assumes all other factors are held constant. The following subsections describe the variables in the regressions.
**Table 4. Associations of Student Characteristics on Subsequent Test Scores**

<table>
<thead>
<tr>
<th></th>
<th>Listening</th>
<th>Speaking</th>
<th>Reading</th>
<th>Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Starting Grade</strong></td>
<td>0.041</td>
<td>0.0894</td>
<td>0.074</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>0.0045</td>
<td>0.0042</td>
<td>0.0046</td>
<td>0.0048</td>
</tr>
<tr>
<td><strong>Square of Starting Grade</strong></td>
<td>-0.004</td>
<td>-0.009</td>
<td>-0.008</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>0.0005</td>
<td>0.0005</td>
<td>0.0005</td>
<td>0.0005</td>
</tr>
<tr>
<td><strong>Initial Subtest Score</strong></td>
<td>0.378</td>
<td>0.397</td>
<td>0.464</td>
<td>0.413</td>
</tr>
<tr>
<td></td>
<td>0.0039</td>
<td>0.0031</td>
<td>0.004</td>
<td>0.0042</td>
</tr>
<tr>
<td><strong>Hmong-Speaking</strong></td>
<td>-0.187</td>
<td>-0.195</td>
<td>-0.13</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>0.0135</td>
<td>0.0123</td>
<td>0.0137</td>
<td>0.014</td>
</tr>
<tr>
<td><strong>Spanish-Speaking</strong></td>
<td>-0.15</td>
<td>-0.044</td>
<td>-0.143</td>
<td>-0.201</td>
</tr>
<tr>
<td></td>
<td>0.0097</td>
<td>0.0089</td>
<td>0.0099</td>
<td>0.0102</td>
</tr>
<tr>
<td><strong>Low Income</strong></td>
<td>-0.197</td>
<td>-0.143</td>
<td>-0.2</td>
<td>-0.206</td>
</tr>
<tr>
<td></td>
<td>0.0094</td>
<td>0.0085</td>
<td>0.0095</td>
<td>0.0097</td>
</tr>
<tr>
<td><strong>IEP</strong></td>
<td>-0.4</td>
<td>-0.321</td>
<td>-0.433</td>
<td>-0.505</td>
</tr>
<tr>
<td></td>
<td>0.0167</td>
<td>0.015</td>
<td>0.0169</td>
<td>0.0174</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>0.374</td>
<td>0.257</td>
<td>0.387</td>
<td>0.439</td>
</tr>
<tr>
<td></td>
<td>0.0096</td>
<td>0.0088</td>
<td>0.0098</td>
<td>0.0101</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>32,243</td>
<td>32,274</td>
<td>32,244</td>
<td>32,223</td>
</tr>
<tr>
<td><strong>R-squared</strong></td>
<td>0.293</td>
<td>0.392</td>
<td>0.364</td>
<td>0.328</td>
</tr>
</tbody>
</table>

Standard errors are below the coefficients; all p values are p<0.001

Source: Authors, using Wisconsin DPI ELL student dataset, 2005-06 through 2013-14.

**Language Groups**

Hmong- and Spanish-speaking students score lower in all four subtests than students speaking other native languages. Hmong-speaking students score lower than Spanish speaking and other ELLs on the listening and reading subtests, while Spanish-speaking students score lower on the reading and writing subtests.

**Low Income Students**

The low income variable is relatively consistent across the four subtest variables. Consistent with prior work, we found that students receiving free or reduced-price lunch perform less well than other students on these exams. The effect size is between 0.14 and 0.2 standard deviations on the four ACCESS test components.

**Disability Status**

Model 1 predicts ELL students with individual education plans will perform significantly worse than ELL students without these plans outlining how schools serve each student with a disability. Listening scores are predicted to be 0.4 standard deviations lower, while speaking scores are 0.3 standard deviations lower. Reading scores are more than 0.4 standard deviations lower, and writing scores are more than 0.5 standard deviations lower.
Key Independent Variables: Age and Initial Proficiency

The following sub-sections are divided by the ACCESS exam component, describing the starting age and initial subtest score. The starting age variable is a quadratic term to capture the complexity of the effect, and therefore the findings discuss curved linear effects.

The results suggest that there is indeed a curvilinear relationship between starting grade and ACCESS scores. Younger and older students perform less well. Having strong initial proficiency is positively related to ACCESS performance—the greater English language skills students have to begin with, the better they perform later. The results are consistent across the different test components in terms of statistical significance and direction, but vary somewhat in terms of effect sizes, which are reported below in more detail.

Listening: The starting age variable in the linear regression has a small, positive coefficient. This coefficient suggests that the older a student entering an ELL program, the more likely the student will perform a few points better on the test. The curvilinear model is more descriptive in interpreting the effects of age. The direction of the starting age coefficient aligns with other research (Hakuta, Butler, & Witt 2000). In the curvilinear model for listening, the effect of starting grade increases until fifth grade, and then begins to decrease. Figure 4 shows the grade coefficients for all of the test components.

At its highest point, grade correlates to almost 10 percent of a standard deviation increase in the future test. Therefore, a fifth-grader entering the ELL program is predicted to perform almost 10 percent of an effect size better on the later test than a student entering ELL in kindergarten.

The initial proficiency term for listening is a linear term. Table 5 shows the correlated effects of an increase in standard deviations, in initial proficiency, on each of the components of the ACCESS exam.

Figure 4. Grade Effect on ACCESS Subtest Scores

Source: Authors, using Wisconsin DPI ELL student dataset, 2005-06 through 2013-14.
As shown in Table 5, the effect size for a one standard deviation increase in initial listening proficiency is 0.378. At half of a standard deviation increase in initial proficiency, the effect size is 0.189, or a 19 percent of one standard deviation improvement in the future test.

Speaking: The starting age variable in the linear regression for speaking has a positive coefficient of 1 percent of a standard deviation. Here the curvilinear regression is also more descriptive in interpreting the effects of age. In the curvilinear model for speaking, the effect of starting grade peaks again at 5th grade. At its peak, the grade in which a student starts an ELL program is responsible for over 22 percent of a standard deviation increase in the future test. A fifth-grader entering the ELL program is predicted to perform more than 20 percent of an effect size higher on the later test than a student entering ELL in fourth grade. The effect size of starting grade for speaking has the largest effect size across the different test components. (See Figure 4 for a graph of the predicted grade effect).

As shown in Table 5, one standard deviation change in initial speaking proficiency is associated with a 0.397 change in the future test score, holding all else constant. Given an increase of half a standard deviation in initial proficiency, we predict an almost 20 percent of a standard deviation increase in the later test score.

Reading: The linear regression for reading predicts starting age to have a positive effect of under 1 percent of a standard deviation. However, the curvilinear regression peaks at fifth grade with an effect size of 18 percent of a standard deviation.

A one standard deviation in initial reading proficiency is associated with a 0.464 of a standard deviation change in the future test. This coefficient is the largest effect size of all the initial proficiency variables. Therefore, a student with a half a standard deviation increase in her/his initial proficiency score is predicted to see 23 percent of an effect size change in the future test.
Writing: The linear regression for writing has a positive effect size of less than 1 percent of a standard deviation. In comparison, the curvilinear regression peaks with the effects of starting grade at fourth grade, with an effect size of 6 percent. The effect of grade on writing is the smallest of the test components. Additionally, the effect of starting age on the future test score becomes negative after eighth grade. This finding highlights the disadvantage older starting ELL students have in learning the writing component of English.

The effect size for initial proficiency in writing is 0.413, or a 41 percent of a standard deviation increase in the future test score for each full deviation increase in initial proficiency. At the half a standard deviation increase in initial proficiency, holding all else constant, we still predict a greater than 20 percent of a standard deviation increase in the future score.

Robustness Tests
We conducted sensitivity analysis to address three concerns. The results are broadly robust to these alternative specifications. First, we ran models to ensure the results are equivalent if we want the models with scale scores for both the initial proficiency and exam scores rather than standardized scores. Second, we considered multiple forms in this model. We began with linear models, and then progressed to quadratic and logit variations.

Finally, we conducted this model using school district fixed effects. Fixed effects capture unobserved differences between school districts, which might bias the results from our findings. When including school district fixed effects, we found no significant changes to our results. A small number of districts were missing appropriate identifiers, which lowered the N in the analysis, and so we chose to include the version of the model with the broader population.

Results on Success in Reaching English Proficiency and on Duration
For our second analysis, we built two models. The first, the proficiency model, examines which factors are associated with an ELL student reaching English proficiency. Conditional on a student reaching proficiency, the duration model examines the number of years that students spend in ELL programs.

Students who are still progressing to reach proficiency and have not graduated from school are excluded from the proficiency and duration models. Older students entering ELL programs have fewer opportunities to take ACCESS exams and test out, compared to students who enter ELL programs in kindergarten and have 13 years to test out. Our models do not include students who enter ELL programs later because they have less time to complete an ELL program and their inclusion would bias our results.

All students in the sample enter the program before or during sixth grade, and have at most seven years to reach proficiency before they graduate from high school. Within our dataset, 7,155 students reached proficiency within primary and secondary school, and about 93.6 percent of those tested out of the program within seven years.

Table 6 shows the results from the two models.
### Table 6. Results: Proficiency and Duration

<table>
<thead>
<tr>
<th></th>
<th>Stage 1 Test Out</th>
<th>Stage 2 Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td>-1.357***</td>
<td>-0.194***</td>
</tr>
<tr>
<td></td>
<td>0.051</td>
<td>-0.0086</td>
</tr>
<tr>
<td>Initial Proficiency</td>
<td>0.818***</td>
<td>-0.701***</td>
</tr>
<tr>
<td></td>
<td>0.0379</td>
<td>-0.0198</td>
</tr>
<tr>
<td>Hmong</td>
<td>0.1569</td>
<td>0.402***</td>
</tr>
<tr>
<td></td>
<td>0.1684</td>
<td>-0.0559</td>
</tr>
<tr>
<td>Spanish</td>
<td>0.202**</td>
<td>0.212***</td>
</tr>
<tr>
<td></td>
<td>0.096</td>
<td>-0.0395</td>
</tr>
<tr>
<td>Low-Income</td>
<td>-0.443***</td>
<td>0.236***</td>
</tr>
<tr>
<td></td>
<td>0.1012</td>
<td>-0.0373</td>
</tr>
<tr>
<td>IEP</td>
<td>-1.616***</td>
<td>0.355***</td>
</tr>
<tr>
<td></td>
<td>0.1768</td>
<td>-0.106</td>
</tr>
<tr>
<td>Constant</td>
<td>8.475***</td>
<td>5.526***</td>
</tr>
<tr>
<td></td>
<td>0.2878</td>
<td>-0.0385</td>
</tr>
<tr>
<td>Observations</td>
<td>5,823</td>
<td>5,435</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.287</td>
<td></td>
</tr>
</tbody>
</table>

**Standard Errors: *** p<0.01, ** p<0.05**

*Source: Authors, using Wisconsin DPI ELL student dataset, 2005-06 through 2013-14.*

**Proficiency Model**

For the proficiency dependent variable, students who reach proficiency are coded one, and those who have not are coded zero. We include in our logit model many of the variables detailed in our prior analysis: entering grade, language, and whether a student is low income or has a disability requiring an IEP. We also include a composite score of performance on the four ACCESS tests to represent initial proficiency.

The results show that Spanish speaking students have the highest probability of reaching proficiency among the three language categories. The magnitude for Hmong students is not statistically significant, although we see a 22 percent increase in odds of reaching proficiency. As expected, low income students and students with disabilities are less likely to reach proficiency compared to their peers. Being a low income student leads to a 36 percent decrease in odds of reaching proficiency, and being a disabled student means an 80 percent decrease in odds of reaching proficiency.

The results of the regression for the duration model show a negative coefficient, which means, on average, the older an ELL student starts the program, the less likely that he or she will reach proficiency before graduating from high school within seven years. To illustrate, if the starting age increases by one, the odds of reaching proficiency will decrease by 74 percent. If the starting age increases by three, the odds will decrease by 98 percent.
The positive sign of the initial composite standardized score indicates that the higher an ELL student’s initial English proficiency, the more likely he or she will test out the program. For every increase of one standard deviation in initial composite standardized score, we expect to see a 127 percent increase in odds of achieve proficiency.

Duration Model
For the duration model, the population we estimate is ELL students who reach proficiency within seven years and enter the program before or during sixth grade. We use an ordinary least squares regression to estimate the model. The dependent variable, duration, is the number of years that ELLs spend in a program to reach English proficiency; composite means the first composite standardized score of ACCESS test when ELL students enter the program. This sample is limited to those who achieved proficiency, and may not generalize to the experience of students who did not achieve proficiency.

The results show that students entering ELL programs in later grades will take less time to test out compared to students who enter in earlier grades. If we increase the initial grade level by five, an ELL student will reach proficiency almost one year sooner.

Our results show that initial proficiency on ACCESS tests is a strong predictor of how long a student takes to test out of ELL programs. On average, the results suggest that if the initial proficiency composite standardized score increases by one standard deviation, an ELL student will spend about 8.5 fewer months in the program to reach proficiency.

For Hmong- and Spanish-speaking students, the results show that it takes longer for them to reach English proficiency relative to other ELL students. On average, Hmong-speaking students spend almost more months more, and Spanish speaking students almost 2.5 months more in ELL programs compared to students from other language groups. Low income students and students with disabilities also take more time to test out.

Robustness Tests
We also tested for six- and eight-year duration and found consistent results. Another robustness test for the proficiency and duration models is to use composite ACCESS scale scores for initial proficiency levels. This method gives the coefficient more descriptive and interpretative meaning in terms of the effects of each of the independent variables on ELL performance. Without standardizing scores, the results of the proficiency model remain similar. However, the coefficient of grade level becomes positive, which indicates that on average, the older an ELL is when starting a program, the more years he or she will take to test out, holding everything else constant. The duration findings when using scale scores are substantively insignificant. The impacts of age, as discussed further below, warrant further research. The results remained similar when we accounted for district fixed effects in the model.

Discussion
Current AMAO reporting is not adequate for school districts receiving the reports. The reports show district results in limited categories and measure success in terms of meeting Adequate
Yearly Progress. Our analysis provides an overview on how individual student characteristics affect how students fare in ELL programs.

Subtest analysis: Subtest analysis (reading, writing, speaking, and listening) can give school districts a more detailed look at the effectiveness of certain aspects of their ELL programs and tell school districts specific characteristics of different language groups. Our analysis of subtest results showed that subtest scores can provide more detailed information about ELL programs and populations. For example, Spanish-speaking students are on average score better than Hmong-speaking students on the speaking exam, but on average score worse than Hmong students on the writing exam.

Initial English proficiency: Not all students in ELL programs start with the same knowledge of English. Students with higher initial English proficiency score better on later tests, have a higher probability of reaching English proficiency, and require less time to reach English proficiency. Increasing the initial proficiency score by 100 points is associated with spending 1.6 fewer years in an ELL program. These metrics can help school districts plan staffing and intervention levels, something the current AMAO reporting does not allow. While school districts cannot control this variable, it is useful information to help shape expectations. For example, knowing estimates of duration in ELL programs based on initial English proficiency can help school districts estimate how many ELL students they will serve in later years.

Starting ELL later: Our results also suggest that starting later in an ELL program (up to fifth grade, when the trend no longer holds) will result in a modest, but statistically significant increase in test scores. This finding interesting compared to other research results that typically suggest a negative correlation between starting age and later performance. This difference might be due to young students gaining test-taking skills over time, so those who enter the ELL program later might have advantages in test-taking relative to students who enter an ELL program at an earlier grade. Our results also show that older students finish ELL programs more quickly, perhaps because they are re-entering a program after changing school districts or having academic difficulties. Some students not initially categorized as ELL who then become designated as ELLs may have relatively higher English skills than average ELL students. Additionally, the scope of our analysis does not include those students who begin ELL programs in high school. Because of these anomalies, using our results beyond ninth grade is not recommended.

While school districts cannot control starting age, it is still important to realize the impacts of this variable. District ELL program expectations should reflect the population’s starting age. Understanding these findings can help school districts evaluate where they want to focus resources.

Language groups: Subgroup analysis can provide a breakdown of language groups and their challenges. Hmong- and Spanish-speaking students score lower in all four categories compared to other language groups. Our analysis suggests that Hmong-speaking students struggle more than Spanish-speaking students on listening and speaking, but do better than Spanish speaking students on reading and writing exams. These findings are particularly applicable to Wisconsin schools, as Spanish- and Hmong-speaking students make up the majority of Wisconsin ELL students. School districts cannot control language groups; however, understanding the strengths
and weaknesses of language groups can help school districts address the diverse needs of their ELL population.

**Poverty and disability:** Consistent with prior research we found that poorer and disabled students fare less well than other students on ACCESS exams. These students have a lower probability of reaching English proficiency and are expected spend more time in ELL programs. This finding should be taken into consideration when evaluating a district’s ELL population or program.

**Contributions to the field:** This analysis contributes to the field of evaluating ELL performance in three significant ways. First, this analysis breaks down English proficiency into the four ACCESS subtests of reading, writing, listening, and speaking. Subtest analysis helps to identify trends in the state as well as strengths and weaknesses within districts. This analysis highlights differences in literacy and oral language. Second, this analysis includes subgroup analysis focusing on Spanish- and Hmong-speaking students in a population with high concentrations of both. Third, this analysis is specific to Wisconsin, which makes our findings directly applicable to Wisconsin school districts.

**Limitations**

While our models provide useful information and contribute to the literature in important ways, there are a few limitations to note. Though our dataset contains longitudinal data, we could not evaluate students who entered ELL programs prior to 2005. In addition, DPI has only partial data for students who entered ELL programs in the past few years and are still enrolled in an ELL program.

Although modeling characteristics of the ELL program in each district was beyond the scope of this analysis, it may be a question for future analysis. DPI and school districts could collect more information on district-specific program characteristics, including program type and teacher characteristics. Our models capture 60 percent of the variance in student ACCESS scale scores, meaning that our models do not capture all factors influencing student performance. Other factors beyond the scope of this analysis may include district-level poverty and parental education level. However, we did use district fixed effects as a robustness test to compensate for some of this unobserved data.

**Recommendations**

Our research and data analysis contribute to several recommendations to improve the methods of measuring ELL student progress at state and district levels, and to help DPI and district administrators identify strengths and weaknesses within ELL programs.

**Recommendation 1:** DPI provide reports to school districts using subtest analysis of reading, writing, listening, and speaking ACCESS scores. Subtest reporting may help describe the strengths and weaknesses within a district. These reports may help district administrators make better informed decisions about how they structure ELL programs and how to best use resources to improve student performance.

**Recommendation 2:** DPI provide reports to school districts using subgroup analysis of the districts’ ELL student populations where appropriate. Subgroup analysis may be inappropriate in
districts with small ELL populations, as the subgrouping might single out students. However, in
districts with significant ELL populations, subgroup analysis may provide more useful
information to district administrators. For example, reporting scores by language group may help
distinguish whether a district has lower achievement for one language group and higher
achievement for another language group.

**Recommendation 3: DPI measure student growth percentiles and report them to school districts as well as to parents of ELL students.** Student growth percentiles show how each student’s test score growth ranks among academically similar students (RAND 2012). To illustrate, a student growth percentile of 90 means that the student scored better than 9 percent of students that performed similar on a prior year’s exam. Student growth percentiles may help DPI and school districts to evaluate the effectiveness of ELL programs by creating a metric that allows for comparison across many school districts in the state and the nation. Student growth percentiles may be particularly useful for districts with small ELL populations, because these data allow for the ability to compare small groups of students more easily with similar students across the state and the WIDA Consortium. For more details about SGPs, see Appendix F.

**Recommendation 4: DPI explore the WIDA dashboard and create a short user guide for school districts.** Through a meeting with Gary Cook of the WIDA Consortium, we learned that WIDA is implementing a new dashboard in 2015. The WIDA dashboard is an online interface that allows WIDA members, including DPI and district administrators, to access visual representations of ELL statistics. In our meeting, we briefly viewed the structure of the dashboard, and the types of information that will be available to DPI and district administrators. We believe that the dashboard presents valuable information in many useful formats. However, district administrators may not have the time or resources to fully explore the dashboard, which may make finding and correctly interpreting information on the dashboard difficult.

In recognition of this challenge, we suggest that DPI develop resources to help districts find and understand information on the dashboard. These resources may include a short user instruction manual explaining to district administrators how to find and interpret the most pertinent information. This brief user instruction sheet could include simplified directions to help district administrators find key graphs, charts, and other useful information on the dashboard. These directions could also explain briefly how to interpret these key visuals. Any instructions provided to school districts should be very brief, easily understandable, and should facilitate quick and easy access to the most pertinent information.

**Conclusion**

Starting age and initial English proficiency have significant effects on a students’ language achievement, measured in terms of completing an ELL program. Current AMAO reporting methods do provide adequate information about ELL programs to identify strengths and weaknesses at the district level. In this report, we analyzed data from the DPI Longitudinal Data System to help DPI construct metrics that better explain the performance of ELL students.

We recommend that DPI provide reports to school districts using subtest analysis of reading, writing, listening, and speaking ACCESS scores. DPI should also provide subgroup analysis of ELL student populations to districts with significant ELL populations. To help evaluate the
effectiveness of ELL programs, we recommend that DPI measure student growth percentiles and report them to school districts as well as to ELL students. We also recommend that DPI explore the WIDA dashboard and create a short user explanation for school districts.

Implementing the recommendations in this report will help school districts to understand strengths and weaknesses of their ELL programs, and, in turn, better allocate resources and target help for students to become proficient in English. Continued development in this area will help reduce the achievement gap between ELLs and non-ELL students in Wisconsin.
Appendix A: Dual-Language Programs

“Dual-language program” is the umbrella term for bilingual programs where the goal is for students to become bilingual and biliterate. Dual-language programs use the partner language for at least half of the instructional day in the elementary years. These programs generally start in kindergarten or first grade and extend for at least five years, and many continue into middle school and high school. The information in this appendix is from DPI’s website (Wisconsin DPI 2015d).

The four main types of dual-language programs differing mainly in the student populations served:

Developmental bilingual programs (also referred to as maintenance or late exit bilingual programs) are those that use the first language of the English language learners (ELLs) in instruction to teach content and ensure that ELLs gain proficiency in English. Students are from the target language population; developmental programs are one-way immersion programs in that the population is homogeneous.

Two-way immersion programs are dual language bilingual programs that serve English speakers and speakers of the partner language in the same classroom, with a 50-50 mix of students from each language population. Usually, the speakers of the partner language are English language learners. Both English and the first or native language of the ELL students are used in content and language arts instruction.

Heritage language programs use the non-English language background (heritage language) of the students. All students share the same non-English language, so programs are considered one-way immersion programs. Students may or may not be proficient in the heritage language. That is, the students may be fluent and the program is a developmental or maintenance program (e.g., Spanish for Spanish speakers), or the language is being renewed/reclaimed in the community (e.g., Native American language revitalization programs where the language is not used by all in a community). The target language is the primary language of content instruction and/or language arts instruction.

World language immersion programs are one-way immersion programs that enroll primarily native English speakers who are learning a world language. The goal of these programs may be for students to become bilingual.
Appendix B: AMAO Reports
This appendix explains the AMAO report provided by DPI (Wisconsin 2014a).

Each AMAO report that a school district receives begins with a summary of the AMAO determinations. Districts that receive Title III funds receive a determination. The report has a section labeled “Did Your District Meet Annual Measurable Achievement Objectives” and provides how many years that district has missed AMAOs, if applicable. The next section of the AMAO report provides a more detailed progression in English language acquisition. This page provides the percentage of English learners achieving the target improvement in composite English language proficiency (e.g., 2013-14 target of 41 percent). The district or consortium makes this determination, which varies on several factors, including the number of ELL students.

Page 3 of the AMAO report consists of a graph that shows the percent change over time of the students who demonstrate the appropriate improvement in their composite English language proficiency scores on the ACCESS for ELLs exam. Many of Wisconsin’s school districts have 20 or fewer ELL, so the confidence intervals in those districts are significantly larger.

AMAO 2 shows the percentage of English learners reaching English language proficiency (e.g., 2013-14 target of 11 percent). The district or consortium makes this determination, which varies depending on the number of English learners, consortium membership, and consortium fiscal agent status. AMAO 2 shows change over time, confidence intervals, and targets. Similar to the graph for AMAO 1, districts with fewer ELL students have larger confidence intervals.

AMAO 3 consists of three components shown through tables on the AMAO report. The first component measures ELL proficiency in reading and mathematics by scores on the WKCE exam. The report provides prior-year and future target percentages in both subject areas. AMAO 3 calculates grade range (elementary, middle, and high) proficiency and test participation. The district makes these determinations, rather than the WIDA Consortium. Meeting targets in each of the grade ranges for proficiency, participation, and graduation rate fulfills AMAO 3.

To determine reading and mathematics proficiency rates, a district must meet the current-year or two-year Wisconsin Student Assessment System test proficiency target. For a district to meet the participation target, all grade range groups in the district must meet current-year or three-year Wisconsin Student Assessment System test participation target in reading and mathematics. The target participation target is 95 percent for all subjects and grade ranges.

Districts fulfill graduation rate targets by meeting a four-year or six-year graduation rate target. Districts can fulfill these targets by meeting the rate target with the current year rate, two-year average rate, or by meeting an improvement target. For instance, if a district has a four-year graduation rate of 60 to 80 percent and makes 2 percent or more improvement, it passes that AMAO. A district also meets the AMAO if its four-year or six-year high school graduation rate exceeds 85 percent. This provision presents a problem in that districts with a small sample size see tremendous gains or declines in graduation rate with just a few students. The AMAO reports are denoted as “NA” in this section if the district has fewer than 20 ELLs in a given cohort.

This following is a sample AMAO for a school district is available from DPI’s website (Wisconsin DPI 2015a).
The Purpose of This Report

State and federal education laws hold districts accountable for the outcomes of Limited English Proficient (LEP) students - also known as English Language Learners (ELLs). These accountability measures are called Annual Measurable Achievement Objectives (AMAOs). This report describes each AMAO and presents district AMAO results. For more on ELL accountability see: http://oea.dpi.wi.gov/acct/ellamao.

About AMAOs

The AMAO 1 determination is based on student growth in the acquisition of the English language. The AMAO 2 determination is based on the attainment of English language proficiency. Both AMAO 1 and 2 are measured using the ACCESS for ELLs® assessment. The AMAO 3 determination is based on the district’s ELL proficiency and test participation rates as measured with the Wisconsin Student Assessment System (WSAS) as well as the district’s ELL graduation rate.

Consortia

AMAO 1 and 2 are evaluated at either the district or consortium level. Whether a district or consortium determination is made is dependent on three factors: 1) the number of student records available for the calculation; 2) consortium membership; and 3) consortium fiscal agent status. Only district-level determinations are made for AMAO 3.

Test results for both the current and prior year for 20 ELLs are required to calculate AMAO 1. AMAO 2 requires only a current year test result for 20 ELLs. If the required test results are present for both the current and prior year then the determination is made at the district level for that AMAO. If the required test results are not present in either the current or prior year and the district is in a consortium then the determination is made at the consortium level. If a district is a consortium fiscal agent then the determination is made at the consortium level regardless of the number of ELLs the district has.

The consortium or district on which results are based is indicated on the AMAO detail pages of this report. Districts choose whether to join a consortium and which consortium to join. For more on consortia see: http://esea.dpi.wi.gov/files/esea/pdf/bul_0401.pdf.

Did Your District Meet Annual Measurable Achievement Objectives?

<table>
<thead>
<tr>
<th>SUMMARY DETERMINATIONS</th>
<th>2011-12</th>
<th>2012-13</th>
<th>2013-14</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMAO 1: Progressing in English language acquisition</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AMAO 2: Reaching English language proficiency</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AMAO 3: ELL-Annual Measurable Objectives (AMO)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Status</td>
<td>Satisfactory</td>
<td>Satisfactory</td>
<td>Satisfactory</td>
</tr>
</tbody>
</table>

An AMAO determination of “NA” indicates that the AMAO is not applicable because the cell size requirement was not met. An overall determination of “Not Title III” means that AMAOs are for informational purposes only because the district did not accept Title III funds.

Data on page one of this report are public information. All other data are intended for use within the district and are not for public review. Because this report may contain personally identifiable student information, the report is for internal use only. Distribution within your district must be in accordance with state and federal privacy laws, and local policy.
AMAO 1 Progressing in English Language Acquisition Detail

Group: Example District

2013-14 Target: 41.0% of ELLs gaining 0.4 or more in composite English language proficiency (ELP) level score

The Purpose of this AMAO

AMAO 1 focuses on student progress in learning English. It requires that districts meet a target percentage of ELLs making progress, defined differently for the two ELP assessments: ELLs taking the ACCESS for ELLs® who gain at least 0.4 composite ELL proficiency level count as progressing while ELLs taking the Alternate ACCESS for ELLs® assessment who gain at least one point in scale score count as progressing.

Method

AMAO 1 is calculated two ways: 1) in the current year, and 2) across two years. AMAO 1 is met if the percentage of ELLs making progress meets the target using either calculation.

Steps for the current-year calculation:
1. Count the number of ELLs in the current year that made progress from the prior year.
2. For ELLs who do not have a prior year score but do have a score from two years prior, count the number of ELLs in the current year that made progress from two years prior. ELLs in this group have their score from two years prior subtracted from their current year score and then divided by two to produce average growth across two years.
3. Add together the counts from steps one and two and divide that count by the total number of ELLs. This results in the percentage of ELLs making progress.

Steps for the two-year calculation:
1. Count the number of ELLs in the prior year that made progress from two years prior.
2. Add the number of ELLs making progress from the current year calculations above.
3. Divide the count of ELLs making progress calculated in step two by the total number of ELLs. This results in the percentage of ELLs making progress.

Targets

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NA</td>
<td>35.0%</td>
<td>37.0%</td>
<td>39.0%</td>
<td>41.0%</td>
<td>43.0%</td>
<td>45.0%</td>
<td>47.0%</td>
<td>49.0%</td>
<td>51.0%</td>
<td>53.0%</td>
<td>55.0%</td>
</tr>
</tbody>
</table>

Results

<table>
<thead>
<tr>
<th>Current or Two Year Calculation</th>
<th>Number of ELLs Making Progress</th>
<th>Number of ELLs with Two Composite Scores</th>
<th>Percent Gaining 0.4 Composite ELP Level Score</th>
<th>Percent Gaining 0.4 Composite ELP Level Score Upper Confidence Interval</th>
<th>Target</th>
<th>Met Target</th>
<th>AMAO 1 Determination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Year</td>
<td>42</td>
<td>63</td>
<td>66.7%</td>
<td>77.1%</td>
<td>41.0%</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Two Year</td>
<td>84</td>
<td>126</td>
<td>66.7%</td>
<td>74.3%</td>
<td>41.0%</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

“Too Few” indicates that the target is not applicable because the district had fewer than 20 ELLs in the current year, the prior year or both years. “Yes-CI” means that the target was met only after applying a 95% confidence interval.
AMAO 1 Progressing in English Language Acquisition Detail

Vertical lines in the charts below represent the 95% confidence interval.

**Current Year Percent Progressing**

**Two Year Percent Progressing**

---

Data on page one of this report are public information. All other data are intended for use within the district and are not for public review. Because this report may contain personally identifiable student information, the report is for internal use only. Distribution within your district must be in accordance with state and federal privacy laws, and local policy.
AMAO 2 Reaching English Language Proficiency Detail

Group: Example District
2013-14 Target: 11.0% of ELLs reaching English language proficiency (ELP)

The Purpose of this AMAO
AMAO 2 focuses on English language proficiency. It requires that districts meet a target percentage of ELLs achieving proficiency as measured by the ACCESS for ELLs® assessment. ELLs who score a composite ELP level score of five or above are considered proficient.

Method
AMAO 2 is calculated two ways: 1) in the current year, and 2) across two years. If the percentage of ELLs achieving English language proficiency meets the target using either calculation then AMAO 2 is met.

Steps for the current year calculation:
1. Count the number of ELLs scoring proficient.
2. Divide by the number of ELLs enrolled. This results in the percentage of ELLs scoring proficient.

Steps for the two year calculation:
1. Count the number of ELLs scoring proficient across two years.
2. Divide by the number of ELLs enrolled across two years. This results in the percentage of ELLs scoring proficient.

If a student is enrolled in both the current year and prior year their test scores from both years are included in the calculation. ELLs taking the Alternate ACCESS for ELLs® assessment cannot score proficient but are included in the number enrolled.

Targets

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.0%</td>
<td>6.5%</td>
<td>8.0%</td>
<td>9.5%</td>
<td>11.0%</td>
<td>12.5%</td>
<td>14.0%</td>
<td>15.5%</td>
<td>17.0%</td>
<td>18.5%</td>
<td>20.0%</td>
<td>NA</td>
</tr>
</tbody>
</table>

Results

<table>
<thead>
<tr>
<th>Current or Two Year Calculation</th>
<th>Number Reaching English Language Proficiency</th>
<th>Percent Reaching English Language Proficiency</th>
<th>Percent Reaching English Language Proficiency Upper Confidence Interval</th>
<th>Target</th>
<th>Met Target</th>
<th>AMAO 2 Determination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Year</td>
<td>22</td>
<td>78</td>
<td>28.2%</td>
<td>39.0%</td>
<td>11.0%</td>
<td>Yes</td>
</tr>
<tr>
<td>Two Year</td>
<td>53</td>
<td>165</td>
<td>32.1%</td>
<td>39.6%</td>
<td>11.0%</td>
<td>Yes</td>
</tr>
</tbody>
</table>

“Too Few” indicates that the target is not applicable because the district had fewer than 20 ELLs in the current year, the prior year or both years. “Yes-CI” means that the target was met only after applying a 95% confidence interval.
AMAO 2 Reaching English Language Proficiency Detail

Vertical lines in the charts below represent the 95% confidence interval.

**Current Year Percent Proficient**

**Two Year Percent Proficient**

Data on page one of this report are public information. All other data are intended for use within the district and are not for public review. Because this report may contain personally identifiable student information, the report is for internal use only. Distribution within your district must be in accordance with state and federal privacy laws, and local policy.
A MAO 3 Annual Measurable Objectives (AMOs) for ELLs Detail

Group: Example District
2013-14 Target: Multiple targets, see below.

The Purpose of this AMO
A MAO 3 is made up of three components: 1) ELL proficiency in reading and mathematics as measured by the Wisconsin Knowledge and Concepts Examination (WKCE) and the Wisconsin Alternate Assessment for Students with Disabilities (WAA-SwD); 2) ELL test participation in reading and mathematics on the WKCE and WAA-SwD; and 3) ELL high school graduation.

A MAO 3 is calculated at each grade range (elementary, middle, and high) for both the proficiency and test participation components. Graduation is calculated for high school only. If targets are met in each of the grade ranges for proficiency and participation and the graduation target is met then A MAO 3 is met.

Method for Calculating Reading and Mathematics Proficiency Rates
For a district to meet the proficiency target, it must meet either the current-year or two-year Wisconsin Student Assessment System (WSAS) test proficiency target. Proficiency targets for A MAO 3 are the same as those for AMOs: http://oea.dpi.wi.gov/acct/amo.

Targets

<table>
<thead>
<tr>
<th>Subject</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>9.6%</td>
<td>16.3%</td>
<td>23.0%</td>
<td>29.7%</td>
<td>36.4%</td>
<td>43.1%</td>
<td>49.8%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>24.0%</td>
<td>30.9%</td>
<td>37.8%</td>
<td>44.7%</td>
<td>51.6%</td>
<td>58.5%</td>
<td>65.4%</td>
</tr>
</tbody>
</table>

Results

<table>
<thead>
<tr>
<th>Subject</th>
<th>Grade Range</th>
<th>Current Year ELLs Proficient</th>
<th>Current Year ELLs Tested</th>
<th>Current Year Percent Proficient</th>
<th>Two Year ELLs Proficient</th>
<th>Two Year ELLs Tested</th>
<th>Two Year Percent Proficient</th>
<th>Target</th>
<th>Met Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>Elementary</td>
<td>8</td>
<td>29</td>
<td>27.6%</td>
<td>18</td>
<td>60</td>
<td>30.0%</td>
<td>29.7%</td>
<td>Yes</td>
</tr>
<tr>
<td>Reading</td>
<td>Middle</td>
<td>4</td>
<td>20</td>
<td>20.0%</td>
<td>6</td>
<td>45</td>
<td>13.3%</td>
<td>29.7%</td>
<td>Yes-CI</td>
</tr>
<tr>
<td>Reading</td>
<td>High</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Mathematics</td>
<td>Elementary</td>
<td>14</td>
<td>29</td>
<td>48.3%</td>
<td>32</td>
<td>60</td>
<td>53.4%</td>
<td>44.7%</td>
<td>Yes</td>
</tr>
<tr>
<td>Mathematics</td>
<td>Middle</td>
<td>6</td>
<td>20</td>
<td>30.0%</td>
<td>9</td>
<td>45</td>
<td>20.0%</td>
<td>44.7%</td>
<td>Yes-CI</td>
</tr>
<tr>
<td>Mathematics</td>
<td>High</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

"NA" indicates that the target is not applicable because the grade range had fewer than 20 ELLs in the current year or across the current and prior year combined. "Yes-CI" means that the target was met only after applying a 95% confidence interval.
AMAO 3 Annual Measurable Objectives (AMO) for ELLs Detail

Group: Example District

2013-14 Target: Test 95% of ELLs.

Method for Calculating Reading and Mathematics Test Participation Rates

For a district to meet the participation requirement, all grade range groups in the district must meet either the current-year or three-year Wisconsin Student Assessment System (WSAS) test participation target in both reading and mathematics. The three-year rate is based on two years of data when three years are not available.

Targets
The test participation target is 95% for all subjects and grade ranges.

Results

<table>
<thead>
<tr>
<th>Subject</th>
<th>Grade Range</th>
<th>Current Year ELLs</th>
<th>Current Year Rate</th>
<th>Three Year ELLs</th>
<th>Three Year Rate</th>
<th>Target</th>
<th>Met Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>Elementary</td>
<td>31</td>
<td>100.0%</td>
<td>104</td>
<td>96.2%</td>
<td>95.0%</td>
<td>Yes</td>
</tr>
<tr>
<td>Reading</td>
<td>Middle</td>
<td>26</td>
<td>100.0%</td>
<td>78</td>
<td>100.0%</td>
<td>95.0%</td>
<td>Yes</td>
</tr>
<tr>
<td>Reading</td>
<td>High</td>
<td>9</td>
<td>NA</td>
<td>24</td>
<td>NA</td>
<td>95.0%</td>
<td>NA</td>
</tr>
<tr>
<td>Mathematics</td>
<td>Elementary</td>
<td>31</td>
<td>100.0%</td>
<td>104</td>
<td>100.0%</td>
<td>95.0%</td>
<td>Yes</td>
</tr>
<tr>
<td>Mathematics</td>
<td>Middle</td>
<td>26</td>
<td>100.0%</td>
<td>78</td>
<td>100.0%</td>
<td>95.0%</td>
<td>Yes</td>
</tr>
<tr>
<td>Mathematics</td>
<td>High</td>
<td>9</td>
<td>NA</td>
<td>24</td>
<td>NA</td>
<td>95.0%</td>
<td>NA</td>
</tr>
</tbody>
</table>

"NA" indicates that the target is not applicable because the grade range had fewer than 20 ELLs in the current year or across the current and prior two years combined.
AMAO 3 Annual Measurable Objectives (AMO) for ELLs Detail

Group: Example District  
2013-14 Target: Multiple targets, see below.

Method for Calculating Graduation Rate
For a district to meet the graduation rate target, it must meet either the four-year or six-year graduation rate target. Targets can be met by meeting the graduation rate target with the current year rate, the two-year average rate, or by meeting an improvement target. Graduation rate targets for AMAO 3 are the same as those for AMOs: http://oea.dpi.wi.gov/acct/amo.

Wisconsin uses the graduation rate goal of 85 percent. Separate graduation rate improvement targets are used for the four- and six-year rates. A school meets the graduation rate AMO for graduation if 1) the graduation rate for the most recent year, or for the most recent two years combined, meets the 85 percent goal; or 2) the improvement in graduation rate meets the applicable target.

The district meets the graduation requirement if it:
- has a four-year or six-year graduation rate above 85%;
- has a four-year graduation rate of 60% to 80% and makes 2% or more improvement; or
- has a six-year graduation rate of 60% to 80% and makes 5% or more improvement; or
- has a four-year or six year graduation rate of less than 60% and makes 5% or more improvement.

For more information on graduation in Wisconsin: http://graduation.dpi.wi.gov.

### Targets

<table>
<thead>
<tr>
<th>Graduation Rate</th>
<th>Four-Year Graduation Rate Improvement Target</th>
<th>Six-Year Graduation Rate Improvement Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greater than 85%</td>
<td>No improvement target - met graduation rate target</td>
<td>No improvement target - met graduation rate target</td>
</tr>
<tr>
<td>60% to 85%</td>
<td>2.0%</td>
<td>5.0%</td>
</tr>
<tr>
<td>Less than 60%</td>
<td>5.0%</td>
<td>5.0%</td>
</tr>
</tbody>
</table>

### Results

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Current Year Graduates</th>
<th>Current Year ELLs in Cohort</th>
<th>Current Year Graduation Rate</th>
<th>Two Year Graduates</th>
<th>Two Year ELLs in Cohort</th>
<th>Two Year Graduation Rate</th>
<th>Improvement Target</th>
<th>Improvement Change</th>
<th>Met Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four-Year</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Six-Year</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

“NA” indicates that the target is not applicable because the grade range had fewer than 20 ELLs in the cohort in the current year or across the current and prior year combined.

Data on page one of this report are public information. All other data are intended for use within the district and are not for public review. Because this report may contain personally identifiable student information, the report is for internal use only. Distribution within your district must be in accordance with state and federal privacy laws, and local policy.
Appendix C: Federal Funding

Title III of the No Child Left Behind Act provides for federal financial support for state and local education agencies. DPI receives funding based on the number of ELL students and allocates those funds to school districts. Title III assists school districts with the grant application process, compliance, and legal requirements, Elementary and Secondary Education Act monitoring, accountability issues, and professional development. The following goals receive funding priority:

- Increase English proficiency and academic achievement in core academic subjects taken by students whose proficiency in English is limited by providing high-quality language instruction programs and content area teaching;
- Provide high-quality professional development to enable classroom teachers to deliver effective sheltered content and English language instruction; and
- Develop, implement, and provide extended day, weekend, and summer opportunities for English language and academic content instruction for limited-English proficient students, (Wisconsin DPI 2014b).

Wisconsin school districts may receive categorical aids to reimburse a portion of the approved costs of their bilingual/bicultural programs (Wisc. Stat §115.995 1999). Districts may use these funds for staff salaries and books and materials used in English language instruction.

Federal funding follows the same allocation goals as Wisconsin funds, but districts may use these funds to pay for parental involvement activities and professional development. Data from the 2010-11 school year shows that Title III allocated approximately $123 per pupil. Title III allocated a total of about $5.4 million to public school districts and approximately $195,000 to non-public schools (Wisconsin DPI 2014b).

Additional grant and funding opportunities for school districts include: The Refugee School Impact Grant, National Professional Development Grant, Title I of the Improving America’s Schools Act: Title I, Part A (Improving Basic Programs Operated by Local Educational Agencies) of the No Child Left Behind Act of 2001, Migrant Education: Title I, Part C (Education of Migratory Children) of the Improving America’s Schools Act, and the Immigrant Children and Youth Grant: Title III, No Child Left Behind, English Language Acquisition: Title III-A of No Child Left Behind (Wisconsin DPI 2014b).
Appendix D: English Language Proficiency Levels

Appendix D provides the descriptions of each proficiency level, as defined by Wisconsin (Wisconsin DPI 2015f).

The definitions of the five limited-English language proficiency levels, as well as Level 6, one of two fully-English language proficiency levels, are from PI 13.08(3)(1)-(6), Wisconsin Administrative Rule. Level 7, the other fully-English language proficiency level, is used for purposes of state reporting/state testing.

Level 1—Beginning/Preproduction [WIDA level = Entering]:
A pupil shall be classified level 1 if the pupil does not understand or speak English with the exception of a few isolated words or expressions.

Level 2—Beginning/Production [WIDA level = Beginning]:
A pupil shall be classified level 2 if all of the following criteria are met:
(a) The pupil understands and speaks conversational and academic English with hesitancy and difficulty.
(b) The pupil understands parts of lessons and simple directions.
(c) The pupil is at a pre-emergent or emergent level of reading and writing in English, significantly below grade level.

Level 3—Intermediate [WIDA level = Developing]:
A pupil shall be classified level 3 if all of the following criteria are met:
(a) The pupil understands and speaks conversational and academic English with decreasing hesitancy and difficulty.
(b) The pupil is post-emergent, developing reading comprehension and writing skills in English.
(c) The pupil’s English literacy skills allow the student to demonstrate academic knowledge in content areas with assistance.

Level 4—Advanced Intermediate [WIDA level = Expanding]:
A pupil shall be classified level 4 if all of the following criteria are met:
(a) The pupil understands and speaks conversational English without apparent difficulty, but understands and speaks academic English with some hesitancy.
(b) The pupil continues to acquire reading and writing skills in content areas needed to achieve grade level expectations with assistance.

Level 5—Advanced [WIDA level = Bridging]:
A pupil shall be classified level 5 if all of the following criteria are met:
(a) The pupil understands and speaks conversational and academic English well.
(b) The pupil is near proficient in reading, writing, and content area skills needed to meet grade level expectations.
(c) The pupil requires occasional support.

Level 6—Formerly Limited-English Proficient/Now Fully-English Proficient:
A pupil shall be classified level 6 if all of the following criteria are met:
(a) The pupil was formerly limited-English proficient and is now fully English proficient.
(b) The pupil reads, writes, speaks and comprehends English within academic classroom settings.

Level 7—Fully-English Proficient/Never Limited-English Proficient:
The student was never classified as limited-English proficient and does not fit the definition of a limited-English proficient student outlined in either state of federal law.
# Appendix E: Uses of Raw, Scale, and Proficiency Level Scores

This table is from the “Access for ELLs: Interpretive Guide for Score Reports” (WIDA 2014a, 8).

<table>
<thead>
<tr>
<th>Information provided &amp; Suggested uses</th>
<th>Inappropriate/Incorrect conclusions &amp; uses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Raw Scores</strong></td>
<td></td>
</tr>
<tr>
<td>□ Provides the number of items the student answered correctly out of the total number of items</td>
<td>□ Generalizations about student performance such as academic content knowledge or classroom achievement</td>
</tr>
<tr>
<td>□ Provides a glimpse into how a student performs by language domain by ELD Standard</td>
<td>□ Tracking student progress between school years</td>
</tr>
<tr>
<td>□ Provides some information about where language instruction can improve</td>
<td>□ Comparing results with other students</td>
</tr>
<tr>
<td>□ Only provided on the Teacher Report</td>
<td>□ Does not represent item difficulty levels</td>
</tr>
<tr>
<td><strong>Scale Scores</strong></td>
<td></td>
</tr>
<tr>
<td>□ Provides a psychometrically derived score (accounting for all tier and grade level differences) for each language domain (Listening, Speaking, Reading, and Writing)</td>
<td>□ Comparisons cannot be made across Listening, Speaking, Reading, and Writing domains only within domains</td>
</tr>
<tr>
<td>□ Scores are reflected in a scale from 100-600</td>
<td></td>
</tr>
<tr>
<td>□ Monitor Student growth over time (within a language domain, using growth charts)</td>
<td></td>
</tr>
<tr>
<td>□ Provided on the Teacher &amp; Student Roster Reports</td>
<td></td>
</tr>
<tr>
<td><strong>Proficiency Level Scores</strong></td>
<td></td>
</tr>
<tr>
<td>□ Provides a score in terms of the six WIDA language proficiency levels</td>
<td>□ Provide one source of data and should be used in conjunction with other data sources to making decisions about instruction, assessment and services for English Language Learners.</td>
</tr>
<tr>
<td>□ Provides individual domain scores which can be used with the WIDA CAN DO Descriptors to get a realistic profile of the student’s English language performance</td>
<td>□ To monitor growth over time, it is recommended to use scale scores and not the PL scores.</td>
</tr>
<tr>
<td>□ Inform targeted language instruction using the WIDA ELD Standards</td>
<td></td>
</tr>
<tr>
<td>□ Provides information to help determine program eligibility</td>
<td></td>
</tr>
<tr>
<td>□ Provided on the Parent/Guardian, Teacher and Student Roster Reports</td>
<td></td>
</tr>
</tbody>
</table>
Appendix F: Student Growth Percentiles

Student growth percentiles show how each student’s test score growth ranks among academically similar students (RAND 2012). Percentile ranks for English proficiency are based on other students who started the previous school year at the same English proficiency level. The middle student growth percentile indicates how well the student performed relative to similar students.

Parents and teachers can often interpret student growth percentiles easily, because they are similar to commonly used percentiles such as growth and weight charts (Betebenner 2009). To illustrate, a student growth percentiles of 90 means that the student scored better than 90 percent of students.

Student growth percentiles may provide useful information to evaluate the effectiveness of ELL programs, particularly in districts with small ELL populations. Student growth percentiles would allow DPI and district administrators to compare small ELL populations with a larger group of similar students. This larger group could include all Wisconsin districts and students from other states that are members of the WIDA Consortium. Demographics and initial English proficiency level could help determine whether the larger group is an appropriate comparison for individual districts.

Although SGPs can be quite useful, there are some limitations. A baseline is required to measure student growth. First-year students have no baseline, and many students may be in an ELL program for a short time, which makes measuring a baseline difficult. Akram, Erickson, and Meyer (2013) note additional issues that practitioners should be concerned about: model specification and test measurement error. Specification error is when an independent variable is correlated with the error term. Test measurement error can occur in student growth percentiles when the conditional percentile changes with measurement error. The relationship between the early scores and later scores can be incorrectly estimated if variables that belong to the model are missing. If test measurement error is ignored, incorrect conditional percentiles are estimated, which means that conditional percentiles are systematically over- or underestimated (Akram, Erickson, & Meyer 2013).
References

Abedi, Jamal, ed. 2007. English Language Proficiency Assessment in the Nation: Current Status and Future Practice. University of California, Davis, School of Education.


